Infotel

Getting Started With Arcsys

Getting Started With Arcsys

4.7.2.0 (November 28, 2014)

Infotel SA - Tour Gallieni - 36, avenue du Général de Gaulle - 93175 Bagnolet Cedex - France
Tel: +33 (0)1 48 97 38 38 - Fax : +33 (0)1 48 97 49 00

http://www.infotel.com

Arcsys

Version : 4.7.2.0 (11-28-2014)

by Equipe Arcsys

This document was written for all persons who wish to obtain information on the Arcsys product.

Copy forbidden without explicit authorization. © 2014 Infotel SA All rights reserved.

This document is part of the Arcsys software package, a product designed and developed by Infotel. All rights are reserved for Infotel.

This document was written for all owners of a valid Arcsys license.

Similarly, readers agree to respect the confidential nature of this document.

Table of Contents

Introduction 1
- Purpose 1
- Reference Documents 1
- Definitions and Abbreviations 1

Understanding the Archiving Market 2
- Background 3
- Solutions 4

Objectives and Principles 5
- Objectives 6
- Basic Principles 7
- System Components 9
- External Services 11

"OCSTIS" Model Coverage 12
- Openness of the System 13
- System Continuity 15
- Scalability in Arcsys 19
- Traceability Mechanisms 21
- Integrity Mechanisms 24
- Digital signature 27
- System Security 28

Technological Features of the Software 31
- Functional Architecture 32
- Physical Location of Components 35
- Arcsys Media Manager (ArcMover) 41
- Asynchronous Tasks 43
- Synchronous Tasks 48

Advantages 50
- Introduction 51
- Advantages for Business Activity 52
- Advantages for the User 53
- Advantages for the IT Department 54
- Compliance with Standards and Legislation 55

A. Appendices 56
- Registered Trademarks 63
- We want to know your opinion! 64

 © Infotel 2014 - Copy forbidden without explicit authorization.

List of Figures

1. OCSTIS Model 7
2. System Components 9
3. Concept of Objects in Arcsys 14
4. Risk of Dependency 15
5. Storage Infrastructure 20
6. Traceability in Arcsys 21
7. Timestamping Request Statuses 22
8. Integrity Management 24
9. Example of a Nearline Storage Policy 29
10. Functional Architecture 32
11. Graphical Interface 34
12. Arcsys Components 35
13. The ArcMover Functions 41
14. Organization of the Manifest File 44
15. Process Flow of the Archiving Request Statuses 45
16. Process Flow of the Restore Request Statuses 46

 © Infotel 2014 - Copy forbidden without explicit authorization.

Introduction

Purpose

This document presents the functions and techniques implemented by the multi-platform archiving and restore product: Arcsys.

Reference Documents

Getting Started With Arcsys: **Arcsys-presentation-4.7.2.0-en.pdf**

Arcsys Installation Guide: **Arcsys-installation-4.7.2.0-en.pdf**

Arcsys Administration Guide: **Arcsys-administration-4.7.2.0-en.pdf**

Arcsys Web Interface User Manual: **Arcsys-web-4.7.2.0-en.pdf**

Arcsys Web Interface End User Manual: **Arcsys-web-end-user-4.7.2.0-en.pdf**

Arcsys Troubleshooting Guide: **Arcsys-error-4.7.2.0-en.pdf**

Arcsys Java API Guide: **Arcsys-api-4.7.2.0-en.pdf**

Arcsys Functional Description Manual: **Arcsys-functional-description-4.7.2.0-en.pdf**

Arcsys Releases Guide: **Arcsys-release-4.7.2.0-en.pdf**

Arcsys Compatibility Guide: **Arcsys-compatibility-matrix-4.7.2.0-en.pdf**

Definitions and Abbreviations

See the Glossary in Appendix.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Understanding the Archiving Market

 © Infotel 2014 - Copy forbidden without explicit authorization.

Background

The increase in volume of IT environments, together with a sharp rise in regulations, has highlighted the need to classify and retain certain types of information in companies. There are a number of areas in business that have a great need to implement a strategic archiving project.

Financial and Technical Areas

These concerns are often the driving forces behind IT production. This category contains HSM (Hierarchical Storage Management) software designed to manage increasing volumes by offering more streamlined storage solutions. This category also contains niche domains as different as: file archiving, database purging, ERP archiving, email archiving and project archiving, for example.

Documentation

Included in this category is archiving originating from digital document management and publishing. This type of archiving is done down the line from the software where objects are often contained in proprietary structures. Storage solutions available are often marketed by the publisher of the software in question.

Legacy

This type of archiving is often managed by archivists. Searching for information often requires a number of specialized criteria; defining a record is thus one of the main objectives of this category. The concepts of retention and continuity are very important as this type of archiving is set up for decades, and possibly centuries.

Fiscal or Regulatory

This category contains information with average retention periods (three to ten years) that require traceability. Volumes are often high and the unitary elements have varying forms. Classification is an important factor, as requests, although rare, are unpredictable and restore of such information can be subject to time restrictions.

Evidentiary Law

This category aims to retain information with a view to supply proof in the event of legal dispute (invoice, mail, contract, for example). Retention can deal with a longer period (up to 30 years), and requires checking the integrity.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Solutions

Infotel designed the Arcsys product to cover the variety of objects to archive and with multitude of very diverse objectives. The product addresses a number of aspects:

- Legal archiving: Checks the integrity and traceability of content from its creation to restore.
- Long-term archiving, due to an ongoing concern for continuity.
- File archiving, file directory trees, projects, consistent lots, thanks to features based on business applications, object indexing and detailed searches.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Objectives and Principles

 © Infotel 2014 - Copy forbidden without explicit authorization.

Objectives

The Arcsys software package was designed to maintain company information assets for medium or long-term periods. Paradoxically, even though all companies have specially adapted IT equipment, generally speaking, retention functions over time are not properly taken into account, at least not by IT systems. A number of reasons contribute to this:

Technological Breakthroughs

IT systems are constantly changing, leaving in their wake a number of previous technological advances, incompatibilities with hardware, applications and software packages. Changing from one operating system to another, from one application to another, makes a certain amount of data inaccessible. Of course, IT operators will "translate" current information so as to make this data compatible with new systems: they keep the most "visible" part in perfect condition. But all old and historical data, and data that is rarely accessed or not accessed at all, is often put aside. For financial reasons or due to the amount of work involved, it remains in this state in the unspoken hope that it will be simply forgotten.

Company Memory

A significant amount of data in companies still relies on human memory. Phenomena such as the exponential increase in volumes of data processed, the turnover of personnel in companies and mergers/acquisitions reveal that human memory is not infallible and that the company's information assets are in danger. Long-term retention means must be implemented to protect companies from the risk of amnesia.

Risk Prevention

The last element is the most recent and has seen the emergence of major risks that have not been addressed adequately until now. An increase in regulations, audits, the principle of precaution, risk prevention dealing with fraud, counterfeit, lawsuits, etc. means that companies must plan ahead and justify themselves. They must thus "memorize" masses of old data over increasingly longer periods of time, which are incompatible with the lifespan of IT storage equipment.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Basic Principles

The Arcsys software is an archiving infrastructure geared to core businesses used to manage archiving databases in compliance with the **OCSTIS** model, that is:

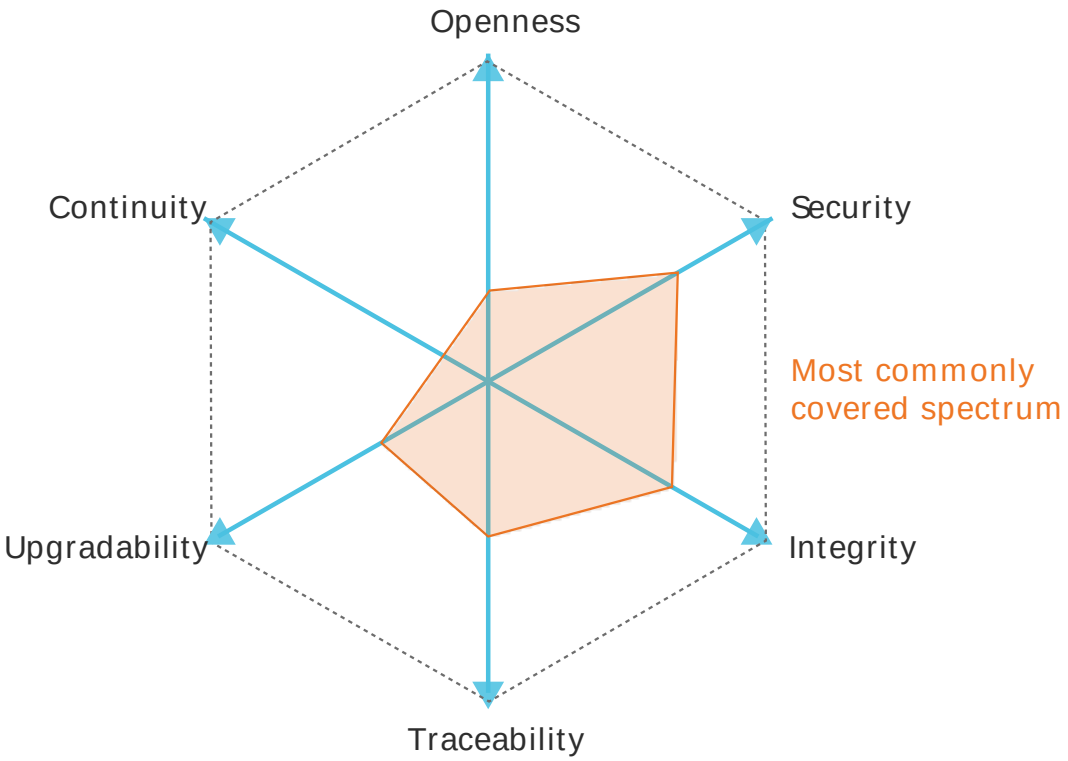

Figure 1. OCSTIS Model

Open

Arcsys accepts all type of items to archive, called objects, without restriction. Basically, Arcsys considers an object to be a series of numerical data (0 and 1) that has a meaning for the issuing application. Arcsys can accept objects sent by different applications concurrently (see Figure 3, “Concept of Objects in Arcsys” [14]).

Continuity

Arcsys was designed to stand up to changes within the limits of current technology. Arcsys is not based on any proprietary elements. It operates on various IT platforms, it does not use a proprietary format at the storage media level, it accepts a variety of storage resources and can switch from one to the other without impacting either the

 © Infotel 2014 - Copy forbidden without explicit authorization.

operator or the applications. Arcsys minimizes the use of proprietary APIs in favor of standard, widely used interface services, which are by nature more durable.

Scalability

Since it is not based on proprietary schemas, Arcsys uses standard techniques with which it can scale according to the IT system of the end client. In this way, Arcsys can operate in "cluster" mode if the client wishes to implement this. Arcsys can even use all the advanced resources the client wishes to put in place in terms of databases. Last but not least, Arcsys can also distribute its processes and therefore leverage the power of multiprocessors.

Traceability

Arcsys "follows" each object to archive step by step throughout the life cycle of the record, from the moment of the archiving request to actual physical archiving. This is also true for each query, restore, migration, request, etc. All these traces are kept in tables in the relational database and in the log files.

Integrity

As soon as it is processed in Arcsys, the record undergoes a mathematical hashing algorithm that calculates a hash for this record. This hash value will be kept as long as the record itself and will be compared to those subject to new calculations at different stages during the record's life cycle (technological migrations, restore, inactivity, for example). Any deliberate or accidental alteration can still be revealed in this way by comparing hashes.

Security

Arcsys uses authentication and authorization means in the company by implementing external interfaces. In this way, you can restrict the scope and investigation field of end users according to their own permissions or the permissions of the groups to which they belong. Arcsys also possesses mechanisms of accountability, non-repudiation and confidentiality.

 © Infotel 2014 - Copy forbidden without explicit authorization.

System Components

Figure 2, “System Components” [9] shows the components of the Arcsys system.

It is important to note that Arcsys software implements generic interfaces with peripheral components as indicated in Figure 2, “System Components” [9]. Priority was given to a solution using proprietary APIs for product continuity and to ensure medium-term scalability.

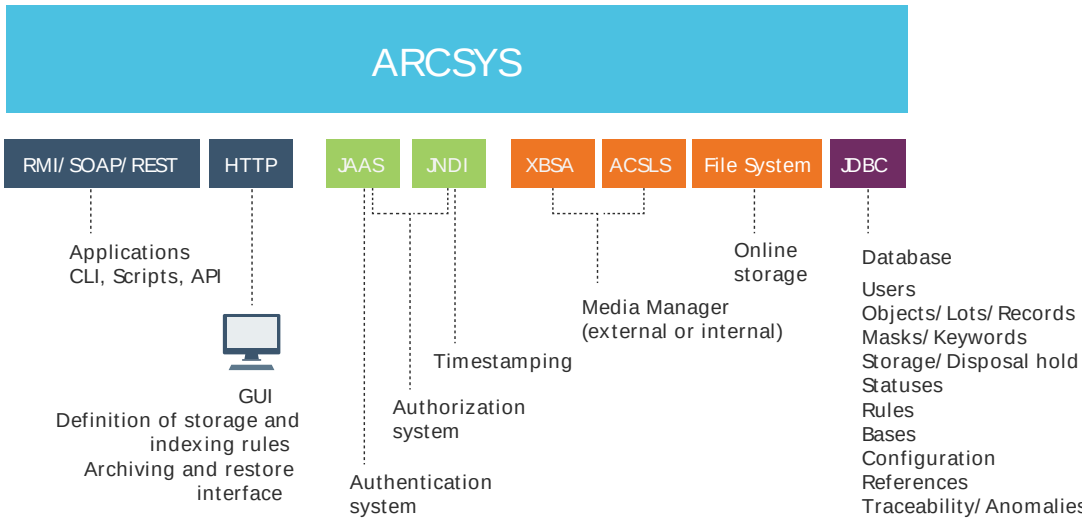

Figure 2. System Components

Central Archiving Engine

Set of synchronous or non-synchronous multi-thread processes:

The engine operates in conjunction with a temporary storage space from which the objects to archive can be transited, either when sent by another agent or when directly transferred to this space.

Relational Database

The relational database contains business metadata relating to records and on which query requests from end users are based. It contains a large amount of additional data relating to indexing rules to be followed, business actions on records (traces), errors encountered, etc. The relational database is the heart of the system.

 © Infotel 2014 - Copy forbidden without explicit authorization.

The Web Agent

This multi-function agent offers both administration options and functions for archiving and restore operations initiated manually by a user.

CLI Command Interface

This is set of commands that can be included directly in the application scripts to call on the archiving system functions to query or restore archives.

API Set

This is a low-level command set, based on RMI or SOAP protocols, with which CLI command lines or special interfaces can be built (thin client, thick client).

A high-level API based on REST Web Services is also available. This is commercial option for the product (ArcBWS).

 © Infotel 2014 - Copy forbidden without explicit authorization.

External Services

There are a number of external services that interface with Arcsys.

- These external authentication and authorization services are accessed via the interfaces available on the Java platform, i.e. JAAS for user authentication and JNDI for looking up authorizations (through directory services).
- The external time server provides an "authentic" external time system so that system timestamping cannot be disputed.
- The communication interface with the media manager gives orders to archive or restore to the storage media; it is also used to dynamically query the relational database of the media manager.

 © Infotel 2014 - Copy forbidden without explicit authorization.

"OCSTIS" Model Coverage

 © Infotel 2014 - Copy forbidden without explicit authorization.

Openness of the System

Various Objectives to Cover

For the most part, applications for archiving currently still run on dedicated platforms. The need for accounting, EDM, mail, fiscal, IT report, invoicing, check imaging, etc. applications has led to (or may lead to) the chaotic multiplication of contiguous, yet all dedicated archiving applications.

Arcsys is structured in object mode. Whether this object is a document, a relational table, a flow or an invoice, Arcsys deploys a pooled generic environment. This openness naturally results in increased performance at infrastructure and productivity levels, but also guarantees scalability.

Concept of Objects

Given the different types of challenges presented by archiving, we have designed a platform that can perform archiving operations for images and documents, as well as relational tables and request results. The concept of object has thus been adopted. The Arcsys product is capable of addressing all structures (series of 0 and 1). To each of these objects, metadata are associated, which are totally dependent on business data. Subsequent searches are performed on these metadata. There are no limitations regarding allowed object types, format types used, and the environments from which these objects originate.

 © Infotel 2014 - Copy forbidden without explicit authorization.

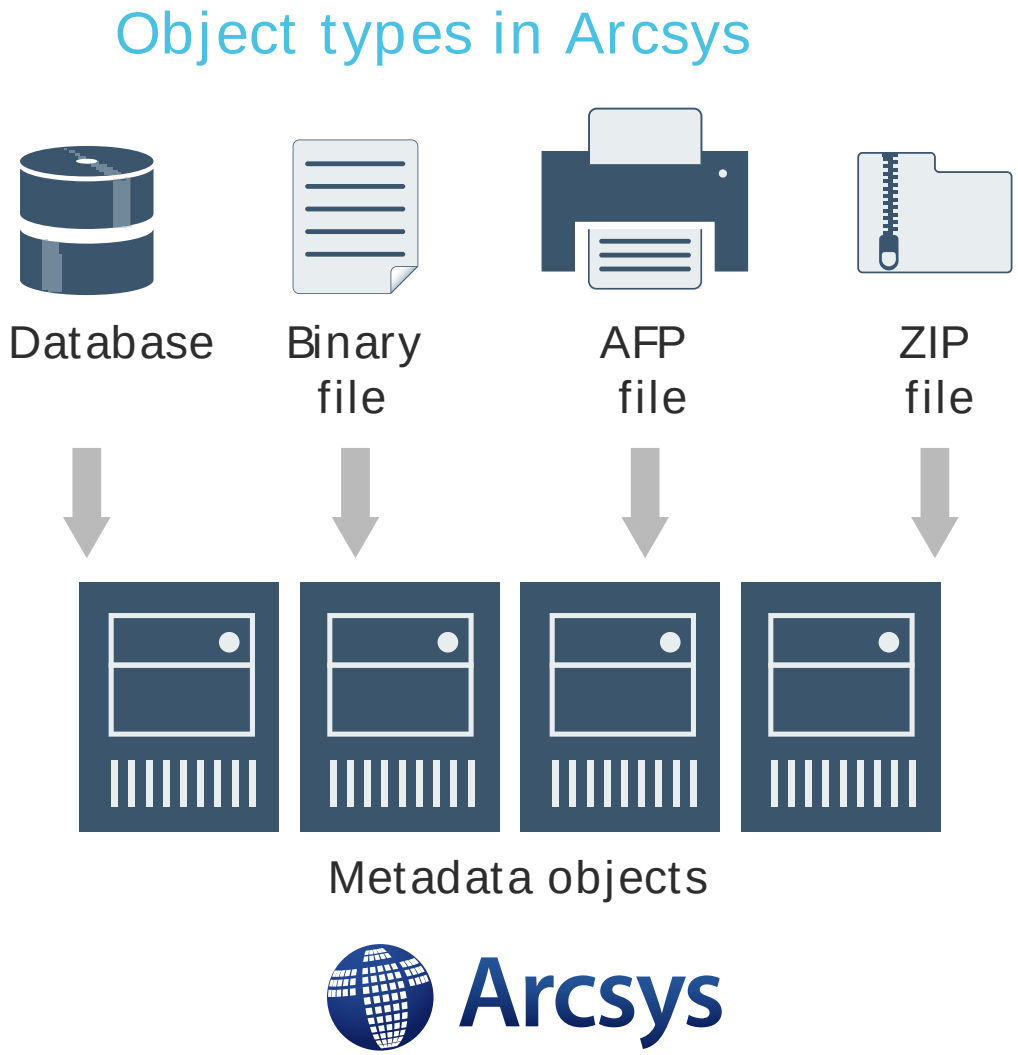

Figure 3. Concept of Objects in Arcsys

The Arcsys application can, if the client so desires, partition applications so that management, requests, and storage of each application can be independent of each other. This is the concept of "base" in Arcsys.

 © Infotel 2014 - Copy forbidden without explicit authorization.

System Continuity

Background

Figure 4, “Risk of Dependency” [15] shows that record continuity depends on three successive layers. The lower layer is the most "visible". Inevitably, it must be upgraded every three to five years to keep up with technological changes.

Without looking to cross the top two layers systematically, nonetheless it must be done and at least, they need to be classified properly, so that the list of formats used that could cause problems is known at all times.

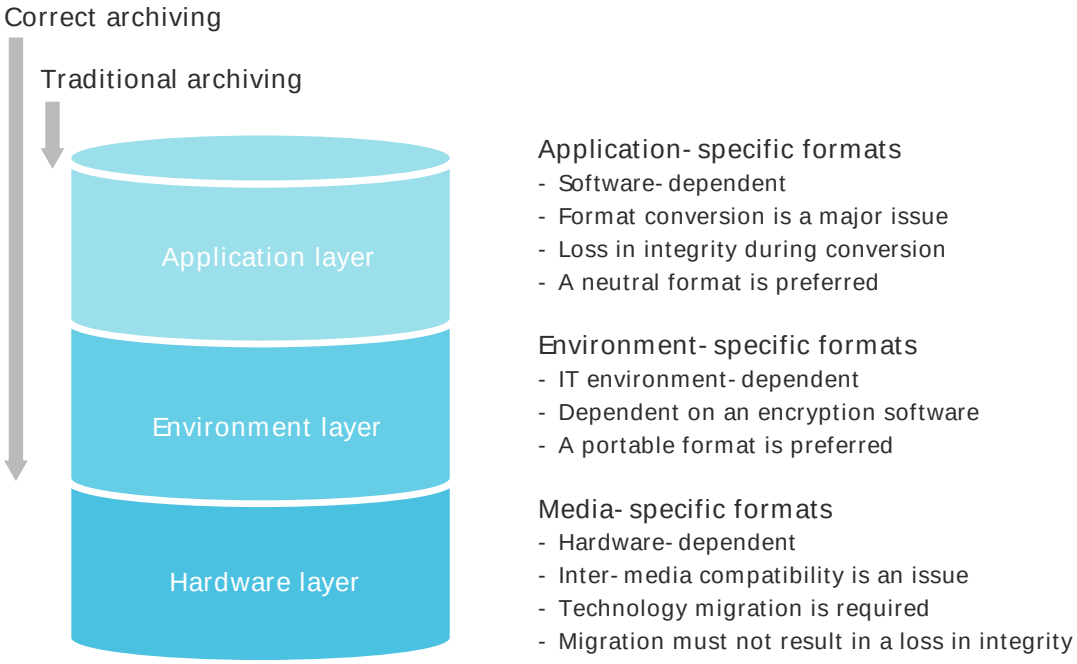
Correct archiving
Traditional archiving
Application layer
Environment layer
Hardware layer
Application- specific formats
- Software- dependent
- Format conversion is a major issue
- Loss in integrity during conversion
- A neutral format is preferred
Environment- specific formats
- IT environment- dependent
- Dependent on an encryption software
- A portable format is preferred
Media- specific formats
- Hardware- dependent
- Inter- media compatibility is an issue
- Technology migration is required
- Migration must not result in a loss in integrity

Figure 4. Risk of Dependency

Arcsys can cross the environment layer and in certain cases, the application layer. However, the latter is the layer that manages presentation; disregarding this presentation could therefore cause problems, particularly for documents.

IT Platforms

Arcsys is based on the most standard products available and always offers an alternative to each component:

- Operating systems of the central platform: z/OS, AIX, Solaris, HP/UX, Windows or Linux

 © Infotel 2014 - Copy forbidden without explicit authorization.

- Relational DBMS :DB2 ®, Oracle ®, SQLserver ® or MySQL
- Online storage systems: None
- Media manager: Internal (ArcMover), or External: NetBackup ® or TSM ®
- Backup tools: None
- Supervision tools: None Arcsys can be monitored using the BMC Patrol ® monitoring tools via use of a third-party KM, for example;
- Authentication: Kerberos (standard on Unix and Windows)
- Authorization tools: LDAPv3 directory

Adaptability of Storage Media

Arcsys accommodates all types of storage media, either supporting a file system-type native interface (online), or supported via the media manager (nearline). Arcsys is independent of all suppliers and all storage technologies.

Media Manager or "Nearline Storage"

Arcsys relies on a media manager to address "nearline" archiving. The advantage is access to the largest possible range of devices on the market without embedding the code in the product for archiving. This simplifies maintenance and ensures continuity of the lower interface, which addresses storage resources.

These choices are based on these two options:

- native writing mode generates records with a neutral format, portable to any existing environment and probably those to come (GNU TAR format). This category includes NetBackup,
- a proprietary format but the editor itself ensures continuity. This category includes TSM.

Arcsys has an XBSA API between the archiving engine and the media manager, which acts as an interface with other products that have not been mentioned previously. XBSA is an open protocol that disregards the type of media manager.

Infotel may consider API implementation for any other media management products, provided that the type writing for this media manager does not involve a proprietary structure and does not infringe on the rules of continuity and porting on all Infotel-imposed environments.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Storage Profile

Arcsys can be used to define a "storage policy" for a record. The storage policy is assigned during archiving. It defines the types of storage media (profiles) that are successfully addressed during the life cycle of the record, the retention periods for each of these media, the number of physical copies of the record, and the logical pool that will be addressed (zones).

When the retention period for a medium has expired, the media manager automatically takes over transfer of the record to the medium according to the hierarchy in compliance with the storage policy. The hashing algorithm (see Integrity Mechanisms) is called before the Arcsys migration routine to check the object hash is always identical to that calculated initially. There is another check after copying on the new medium.

Technological Obsolescence

In the event the company would like to upgrade the storage policy to be in compliance with new needs: extended retention, reduced retention, different storage support, Arcsys will take charge of these new rules allocated by a administrator via the Arcsys graphic interface, and make the change in storage media if necessary.

Compliance with Standards

Arcsys was designed to comply with the most advantageous standards in terms of record retention continuity, that is:

- The ISO standard 14721 or OAIS (Open Archival Information System), designed by an international committee in the scientific community, makes no reference to technology. It defines a reference model for installing archiving systems in terms continuity, functions and interfaces between the sub-components that are ingest, archiving, management and access modules.

- The EAD (Encoded Archival Description) schema is the archival description schema. This presentation and format standard for search instruments is based on a meta-language: XML. It is known as the best implementation tool of the ISAD standard. Its strength lies in its ability to disseminate the archival search instrument and constitute large resource pools.

Continuity of Hashing Algorithms

The chapter Integrity Mechanisms shows the use of hashes for assuring integrity.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Arcsys can use different calculation algorithms for these hashes over time and ensure backward compatibility with hashes used for a number of years previously. For this, the Arcsys indexing mechanisms **are not based on the hash calculated by the algorithm.**

Record Portability

Retaining records for long periods of time presents two problems in the long term:

- A change in storage media;
- A change in the archiving system itself.

Arcsys can perfectly adapt to changes in the storage media, retaining the integrity, and can also inherit the integrity calculated by the previous archiving system and transmit traces during the final migration of integrity, traces, metadata and the manifest to the next archiving system.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Scalability in Arcsys

Redundancy Support

Operation in cluster mode of the relational database is supported natively by the DBMS without impact on Arcsys. In this way, Arcsys supports, for example, Oracle and DB2 clusters.

All other components in the string can be duplicated: Application servers, J2EE application, web server, storage infrastructure.

Scalability in the Number of Processors

Arcsys references the existing engines capable of managing the threads generated by the product's synchronous and asynchronous processes dynamically in an internal table. Arcsys also supports the scalability of 1 to n processors dynamically. Arcsys can distribute the load over all available resources.

Connecting to the Company Network

Arcsys was designed to be able to easily modify the network ports used in order to cross firewalls without restriction.

Arcsys supports TCP/IP networks in standard mode.

Scalability of Storage Media

Arcsys is completely independent of any supplier or storage technology.

Arcsys provides portability from one type of storage to another while guaranteeing record integrity, which provides long-term continuity of the solution.

Storage Infrastructure

Arcsys requires disk space to support the relational database. As its access profile is likely to be highly transactional according to archiving activity and the number of requests, we recommend you use Fibre Channel disks and double access for availability reasons.

At record storage media level, Arcsys clearly distinguishes between two storage profiles:

 © Infotel 2014 - Copy forbidden without explicit authorization.

- Online storage (point (1) in Figure 5, “Storage Infrastructure” [20]) provides quick access to records; this is a disk-type storage with doubled access for availability reasons. Both SATA and Fibre Channel disks are supported. In many cases, SATA disks suffice. Online storage is not mandatory.
- Nearline storage addresses storage media managed by a media manager (points (2)(3)(4) in Figure 5, “Storage Infrastructure” [20]). The differentiating features compared to level (1) is the response time, which will deteriorate due to the intrinsic response time of the media manager. Nearline storage is not mandatory.

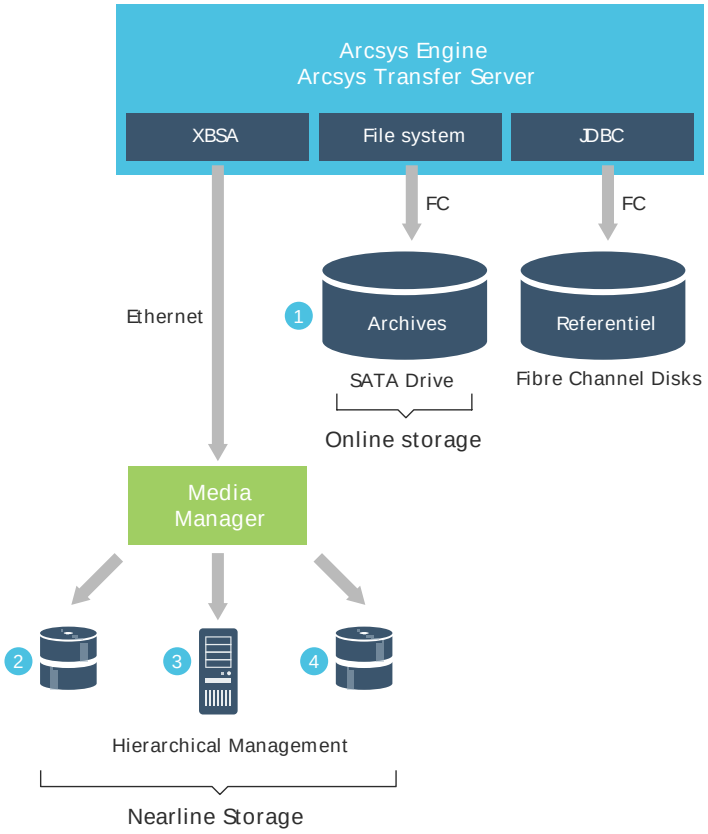

Figure 5. Storage Infrastructure

 © Infotel 2014 - Copy forbidden without explicit authorization.

Traceability Mechanisms

End-to-End Logging

Traceability is handled natively in Arcsys, from request transmission at the application level up to effective writing of the record on the storage medium.

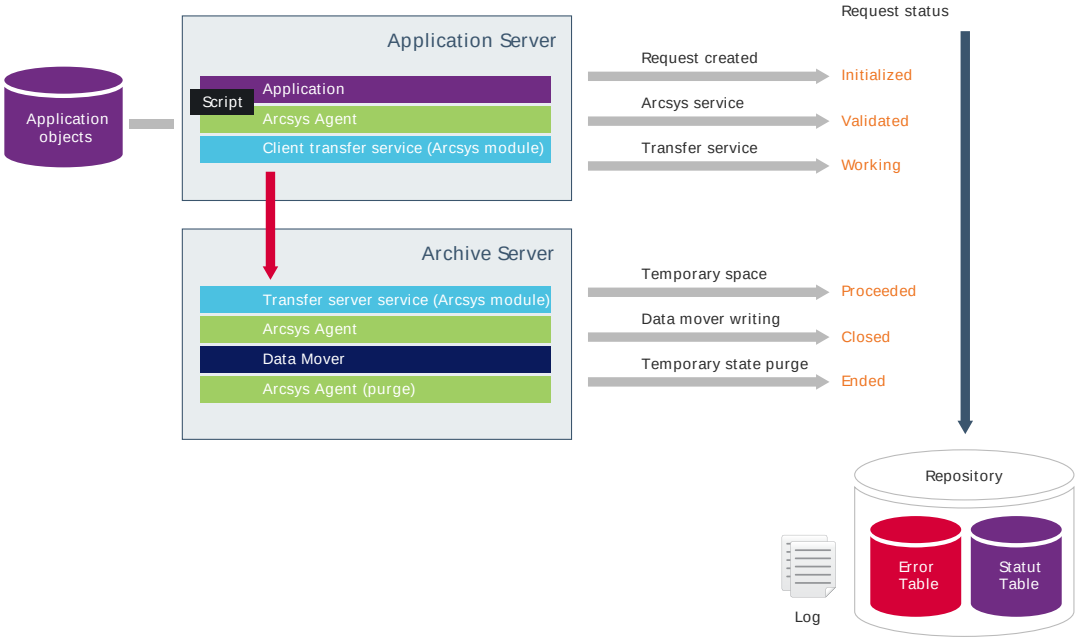

Figure 6. Traceability in Arcsys

 © Infotel 2014 - Copy forbidden without explicit authorization.

Figure 7. Timestamping Request Statuses

Traceability levels

Traceability, in addition to integrity and security, is particularly important in the event of evidentiary law.

All events arising in Arcsys, whether they are linked to configuration (definition of retention and indexing rules, for example) or to daily use (archiving, modification of metadata, restore, migration, etc.) are traced very closely in different forms:

- In the Arcsys relational database;
- In the Arcsys technical log files;
- In the functional trace files activated at collection level.

The log and trace files are themselves archived at time intervals configured in Arcsys. They can be accessed and used by external products such as Patrol.

Arcsys establishes the "Proof Folder" via a module called ArcProofFolder, which is used to generate a complete proof folder associated with document restore.

Non Repudiation and Accountability

Arcsys can capture information at application level and send it directly to the storage platform without the use of intermediaries, making accidental or intentional record substitution, invalidation or addition impossible. Furthermore, the application or user generating the record must sign to interact with Arcsys. The request table provides

 © Infotel 2014 - Copy forbidden without explicit authorization.

the message origin, time and user. The traceability system also mentions each of the intermediate steps: record requested, record archived, record received, record sent for storage, record written with a timestamp and user name, etc. (see Figure 6, “Traceability in Arcsys” [21] and Figure 6).

In this way, a record received by Arcsys cannot be repudiated as the system does not leave room for system fraud during the creation through to restore processes.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Integrity Mechanisms

Unchanged Content

The integrity function of the Arcsys product is based on the call to a mathematical algorithm that calculates a hash for each archived object sent to it. It will last throughout the life cycle of the record. This certificate is dependent only on the data item itself. It is not influenced by the storage medium. The integrity calculation is called a number of times during the life cycle of the data item to attest to the integrity of the lot by comparing it with the original certificate. This hash is stored on the storage medium so that it remains accessible at all times, even if the Arcsys archiving software is not installed.

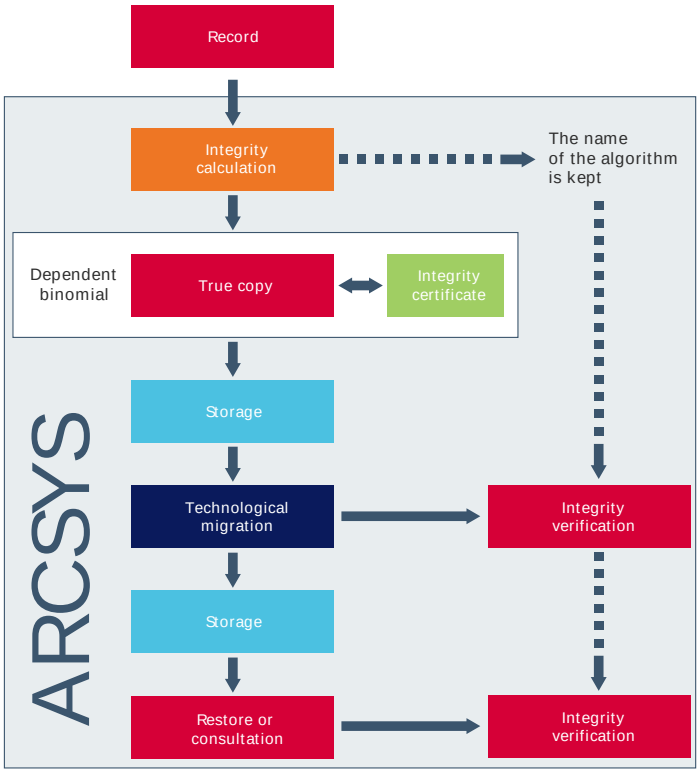

Figure 8. Integrity Management

Note

When the retention period exceeds the life expectancy of the physical media, the concept of global integrity becomes particularly important.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Algorithms Used

Arcsys can use a number of algorithms, including algorithms that may seem sufficient at present (128 bits) but that will quickly become obsolete, as well as more advanced algorithms: 256 and 512 bits. Algorithms are chosen for a given version of Arcsys. Example: One version can be configured with an SHA256 algorithm, and the next version can be configured with SHA512. Backwards compatibility in reading algorithms is, of course, always guaranteed. The name of the algorithm is recorded at the level of the lot description (called manifest, see the definition in the glossary at the end of the document).

To protect records efficiently, algorithms must be upgraded in conjunction with the performance of IT calculation resources. Experts now deem that the SHA-1 protection level no longer offers a sufficiently high security margin and the SHA-256 is becoming increasingly preferred.

This hash calculation module is in particular called by:

- migration routines between media, which, during a change in media resulting from the storage profile rules, calculate the file hash read and compare it with the hash stored on the physical medium;
- and more generally, each time the data item is accessed for viewing, restore or migration operations.

In the event of a consistency error between the hash and the record of origin, Arcsys records the error and tries to access another copy of the record (if the storage profile is such that more than one copy exists).

Checking Integrity Automatically

The hashing calculation routine is also called by internal Arcsys routines for checking physical storage media that periodically reread records stored on physical media and calculate record hash to compare it with the original.

This checking process is triggered periodically. To do this, the engine must be started. This function guarantees that ageing of magnetic media takes place in good conditions.

This function is a commercial option of the Arcsys product.

Non-Rewritability of Records

There are a number of responses to this problem.

 © Infotel 2014 - Copy forbidden without explicit authorization.

The deletion operation is not allowed

Once the record has been created, users can only access it using Arcsys keywords, based on a DBMS.

The records are not directly accessible via a file system, so direct access to delete functions is not allowed.

Arcsys is an archiving system and, as such, does not deliver a standard function for deleting a record. In Arcsys, the records can expire, but they must expire according to the storage policy and after expiry of the scheduled retention. There is no other solution.

Important

In the chapter "Use", and in particular, in a number of screen shots, you can see that the concept of deletion appears. Only an imperatively empty structure can be deleted.

Example: A user creates a collection (or a base, a policy) using the interface and sees a definition error even before record creation. This structure can be deleted if the user has deletion permissions. When a record has been created, users no longer have this permission and the previous points apply.

Restore does not Destroy the Record

Record restore generates a compliant copy of the record, but it does not destroy the record itself on its physical medium.

Re-archiving does not Destroy the Original Record

With the record restored, the re-archiving of this record is technically possible. Nothing can prevent this. From the Arcsys point of view, a re-archived record technically becomes another record because there is at least one difference: the archive date. From a technical point of view, there are thus two physical records that from a traceability point of view are completely different from one another. This is possible because the hash resulting from the hashing algorithm is not used directly as an index to the storage media.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Digital signature

Digital signature mechanisms add an additional level to the integrity aspect. With Arcsys you can attach a digital signature to the document hash. The digital signature is then saved in the relational database, in the XML manifest of the lot and if appropriate in a common signature file.

This function is a commercial option of the Arcsys product (ArcSign).

 © Infotel 2014 - Copy forbidden without explicit authorization.

System Security

Description

The Arcsys engine is accessed in a user session context, i.e. using authentication services and authorization management. Since each company network has these services, Arcsys engine security is based on these external services.

These external authentication and authorization services are accessed via the interface available on the Java platform, i.e. JAAS for user authentication and JNDI for looking up authorizations (through directory services).

Authentication

When the user is identified, a connection to the security management server (Active Directory, Kerberos, PAM, etc.) declared in the configuration files of the product takes place. This operation retrieves "a ticket" that authenticates the end user in a unique way and checks the permissions granted to this user.

Authorizations

The Arcsys interface accesses a mainstream directory service such as Microsoft's Active Directory, which is an LDAP type service, but without limiting the system to this component.

Authorizations from the point of view of functions are handled by the LDAP; authorizations at the level of collections, bases and lots are handled in the Arcsys relational database.

Access and Operating Security

Expiry dates are managed at record level in compliance with the time period requested in the storage policy. A storage policy is a rule (there can be an infinity) that establishes the series of storage media on which a record is to be placed throughout its life cycle. A retention period is associated with each of these media.

 © Infotel 2014 - Copy forbidden without explicit authorization.

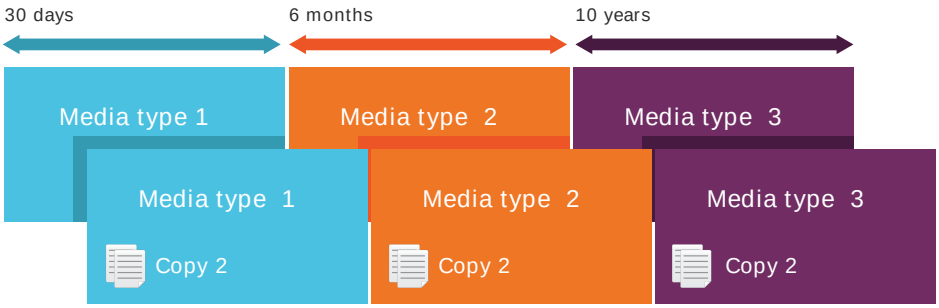

Figure 9. Example of a Nearline Storage Policy

With the example of the Figure 9, “Example of a Nearline Storage Policy” [29], we see a storage profile (exclusively nearline type) in which a succession of three types of media has been defined, designated by logical pools.

In addition, a duplicate copy is requested for each medium.

The expiry date is the sum of the retention periods of n media making up the life cycle, i.e. 10 years and seven months in our example.

Arcsys can generate and manage multiple record copies.

Operational Security

Arcsys can be deployed with a view to a BRP (Business Recovery Plan): each module is duplicated on a number of sites that thus ensure continuity in the event of a technical incident.

Confidentiality

The confidentiality of information during transfer to the network and during intermediate storage on buffer disk zones during the archiving phase is managed either directly by the NFS protocol that takes over the encryption function natively by calling an encryption routine on IP protocol. In the absence of a specific module in the company, Arcsys uses its own encryption module. Confidentiality is configured at the level of each client agent, i.e. certain applications could have their records encrypted during transfer to the network, whereas others would not.

Timestamp Authenticity

To timestamp the activity of the archiving server, Arcsys captures the time system at the level of the server that supports the relational database.

 © Infotel 2014 - Copy forbidden without explicit authorization.

The evidential value of the record will depend on the authenticity of this system time and its traceability. For this, simply connect to a time server that captures the system time on an official source outside the company.

Note

The purpose of this chapter is simply to emphasize to the client company the importance of this point in demonstrating good faith at all times.

This is a prerequisite outside the scope of the Arcsys system that is incumbent on the end client.

External Timestamping

Arcsys provides an option used to perform external timestamping compliant with the RFC3161 standard. Use of this function is recommended in conjunction with the digital signature option. In this way, timestamping is performed on the digital signature.

The complete timestamp response is stored in the Arcsys relational database.

Partitioning Applications

Two complementary utilities manage application partitioning:

- The principle of "bases" that logically partitions records;
- The use of specialized agents per application assigns permissions to users responsible for operating these agents.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Technological Features of the Software

 © Infotel 2014 - Copy forbidden without explicit authorization.

Functional Architecture

Introduction

The diagram below provides an overview of the Arcsys architecture.

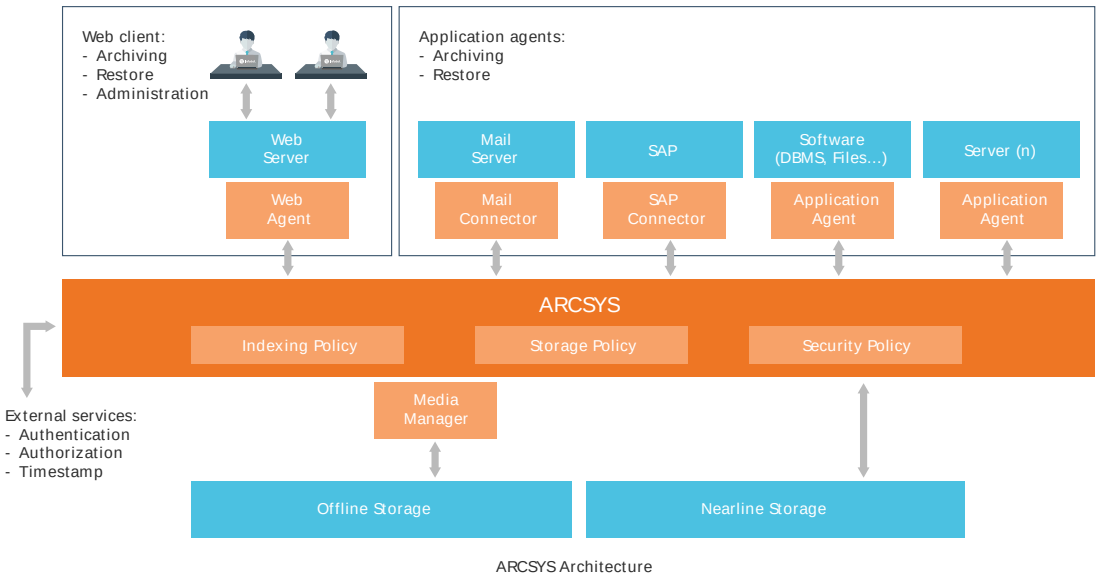

Figure 10. Functional Architecture

It is composed of an engine running synchronous or asynchronous multi-thread processes and client agents.

Application Agent

The application agent is used to archive:

- Data present on an end-user workstation (example: PDF document, etc.) ;
- Data present on an application server, issued from a processing string or extracted by a connector (example: rows extracted from a database, AFP files, documents from an EDM, etc.).

Archiving Engine

The engine operates in conjunction with a temporary storage space from which the objects to archive can be transited, either when sent by another agent or when directly transferred to this space.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Connectors

These are used to extract data from a third-party software package or to bring additional functions to Arcsys.

To date, there are numerous connectors:

- ArcERP for ERP archiving (example: ArcERP for SAP)
- ArcGED for interfacing with EDM software packages (Electronic Document Management). Examples: Nuxeo, FileNet
- ArcMAIL for archiving Exchange or Domino emails
- ArcIP for record injection
- ArcEP for batch record restore
- ArcSIGN for internal and external signature and timestamping management
- ArcAFP for archiving and viewing AFP files
- ArcPAK for managing native ZIP files
- etc.

Please contact Infotel for additional information on these products.

Archiving Agent

The user component is equipped with a web-type graphical interface. It contains functions that are available or not available according to the permissions granted to a group of users to which the connected user belongs.

Generally speaking, there are three main function groups on the agent:

- Administration functions: Manages Arcsys components;
- Operation: Defines archiving policies;
- Business functions: Defines indexing criteria and proceed with archiving.

The web interface is delivered with the properties files in the following two languages: French and English. These files are easily modifiable.

When users identify themselves, a connection to the LDAP server configured takes place thus allowing recovery of all user permissions.

The duration of user sessions can be configured. When the session expires, the user is directed to the identification page.

 © Infotel 2014 - Copy forbidden without explicit authorization.

The figure below gives a general idea of the appearance and usability of the interfaces.

Graphical Interface

On the left, the Menu section is used to find the current function in the Arcsys directory trees and directly access the function in question.

This interface is designed for use with a number of browsers such as Internet Explorer (Windows) or Konqueror (Linux).

Figure 11. Graphical Interface

 © Infotel 2014 - Copy forbidden without explicit authorization.

Physical Location of Components

Introduction

The archiving platform of the Arcsys product is composed of several modules that can be duplicated to absorb any scaling in infrastructure. These modules are located on one or more physical machines according to the capacity loads required for the archiving and restore operations.

Moreover, the general architecture of the Arcsys engine uses a certain number of interfaces. With the exception of the communication interface, these interfaces help to integrate the Arcsys engine in multiple software environments that can be found in an open system.

The figure below illustrates both the components and the flows between these various modules.

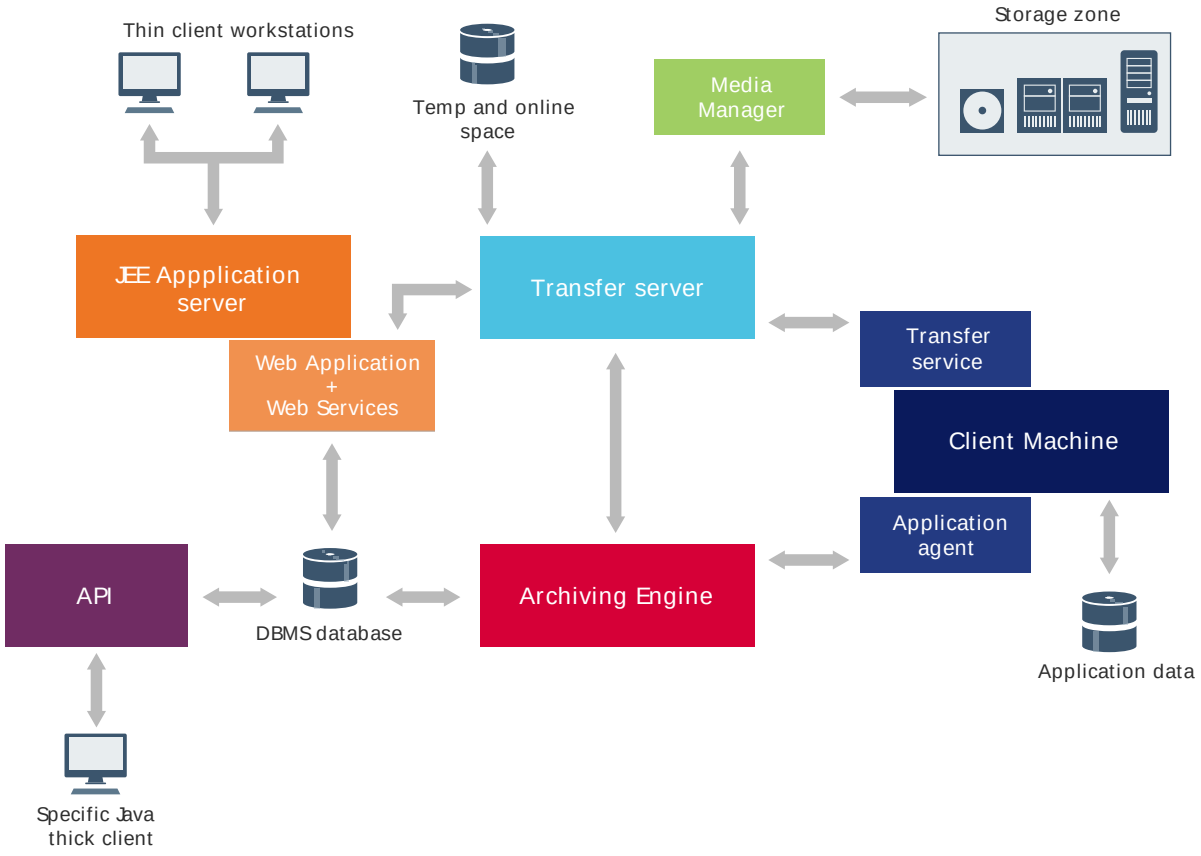

Figure 12. Arcsys Components

 © Infotel 2014 - Copy forbidden without explicit authorization.

Archiving Engines

Each archiving engine is a Java process that uses a number of threads.

Characteristics

This component is used to:

- Carry out archiving and restore requests;
- Prioritize these requests;
- Control the transfer server;
- Control the agents specific to each client application for archiving.

This component:

- Is written in Java;
- Uses the RMI protocol to communicate with the agents;
- Uses the IP sockets to communicate with the transfer server;
- Accesses the relational database via the JDBC interface;
- Checks the identifications and authorizations with JAAS and JNDI.

Transfer Servers

Each transfer server is composed of two processes that are self-monitoring.

Characteristics

This component is used to:

- Control the media manager to archive and restore objects;
- Carry out archiving engine requests;
- Make use of "ONLINE" requests of the application server;
- Retain objects in local with an "ONLINE" retention date;
- Transfer to and from transfer services located on each archiving client machine;
- Manage NFS assembly on archiving client machines.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Transfer Services

The transfer services are presented as two processes controlled by the transfer servers.

Characteristics

This component is used to:

- Transfer from the client machines to the transfer server and vice-versa;
- Update Arcsys software versions on remote machines without manual intervention.

Application Agents

Each application agent is a Java process controlled remotely by an archiving engine.

Characteristics

This component is used to:

- Extract client application objects;
- Build the files of each record;
- Extract the objects of a record to restore them;
- Inject restored objects in the client applications;
- Ensure encryption of transfers;
- Ensure data integrity.

Special Characteristic of the Database Interface (ArcData option)

These agents have a JDBC interface to client databases to capture information in a non-intrusive way in these databases, to prepare the record and direct it to the archiving engine.

The archived information is independent of the internal structure of the base, making long-term restore possible even in the likelihood that the structure of the database has changed.

 © Infotel 2014 - Copy forbidden without explicit authorization.

The extraction agent captures information in a selective manner without disturbing production. The description of the structures and the granularity of the target capture are pre-determined by the graphical interface of the agent.

You can also archive information by identifying it via a date criterion or a criterion given in a specified column.

Furthermore, it is possible to combine information from a database or from a file system within a single record.

Arcsys Web Interface

This is a J2EE application.

Characteristics

This application is used to:

- Present all necessary functions for the product in a "thin" web interface (HTTP or HTTPS protocol);
- Control the transfer server during online display requests of the archived documents.

Relational Database

The relational database is hosted by a relational database manager.

Characteristics

This component is unique on the archiving platform, and:

- Stores all definitions either by business type (indexing policy) or operating type (storage profiles for archived objects);
- Retains all archived object references;
- Facilitates a given object search according to indexing criteria;
- Traces all operations irrespective of their type of archiving, restore, technical work, updates of other components, etc. ;
- Integrated in a standard database manager, it supports the main products on the market (Oracle, DB2, MySQL, SQL Server);
- Uses only functions common to all these managers according to the JDBC protocol.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Modelling

This relational database is simple enough to be easily implemented on the main relational database management systems.

Communication Interface

The engine uses a communication interface based on the Java™ RMI (Remote Method Invocation). This method deploys a part of resource use to the machine hosting the archiving server and not to the machine hosting the client application.

Security Management

The Arcsys engine is accessed in a user session context, i.e. using authentication services and authorization management. Since each company network has these services, Arcsys engine security is based on these external services.

These external authentication and authorization services are accessed via the interfaces available on the Java platform, i.e. JAAS for user authentication and JNDI for looking up authorizations (through directory services).

Authentication

This component implements a Java version of the PAM (Pluggable Authentication Module) model that delivers an authentication service independent of the authentication technologies used.

Authorizations

The external authorization service is accessed via the JNDI interface (Java Naming and Directory Interface). This interface is a Java standard that provides Java applications with a unified interface for directory services. This interface accesses a mainstream directory service such as Microsoft's Active Directory, which is an LDAP type service, but without limiting the system to this component.

Interfaces with the Relational Database

The Arcsys engine requires use of a relational database. Access to this database takes place via an interface for access to data using the JDBC protocol (Java Database

 © Infotel 2014 - Copy forbidden without explicit authorization.

Connectivity) and the connections to this database are distributed via a connection pool.

This connection pool aims to eliminate the impact of initial creation of access to the base, which is costly on performance. The connection pool allows a certain number of pre-established connections to remain permanently available.

The JDBC API allows access to effectively all data sources, database models and flat files. Using this API provides the Arcsys engine with considerable independence in terms of the DBMS chosen to support the relational database.

Arcsys APIs

The APIs can be used in addition to the standard commands or functions available with the product.

Arcsys distinguishes between two levels of APIs:

- Low-level APIs, based on the SOAP or RMI protocols. These APIs are used to manage each type of Arcsys entity (base, collection, lot, storage profile, indexing mask). They access the Arcsys relational database.
- High-level APIs, based on REST (ArcBWS commercial option). They are used to archive, search and restore records.

The APIs are used to access the same options as in the Web Arcsys interface, and are used in particular by all Arcsys connectors to inject or restore records.

A full guide on developing with APIs is available on request.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Arcsys Media Manager (ArcMover)

Overview

In addition to interfaces with NetBackup and TSM products, Arcsys includes its own media manager called ArcMover. This is a commercial option for the product (ArcBWS). This component includes functions that are located on a number of levels:

- at relational database level;
- at archiving engine level;
- at transfer server level;
- at web level;
- via the creation of a new TCP/IP protocol.

ArcMover Functions

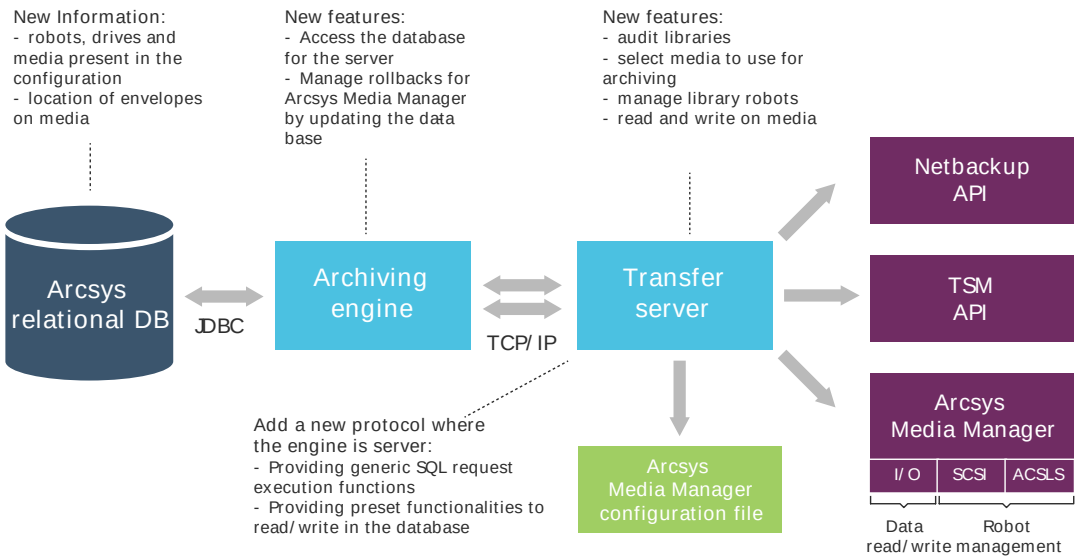
New Information:
- robots, drives and media present in the configuration
- location of envelopes on media
New features:
- Access the database for the server
- Manage rollbacks for Arcsys Media Manager by updating the data base
New features:
- audit libraries
- select media to use for archiving
- manage library robots
- read and write on media
Arcsys relational DB
JDBC
Archiving engine
TCP/ IP
Transfer server
Netbackup API
TSM API
Arcsys Media Manager
I/ O
SCSI
ACSLS
Data read/ write
Robot management
Arcsys Media Manager configuration file
Add a new protocol where the engine is server:
- Providing generic SQL request execution functions
- Providing preset functionalities to read/ write in the database

Figure 13. The ArcMover Functions

Relational Database

Information in the relational database includes in particular:

 © Infotel 2014 - Copy forbidden without explicit authorization.

- Managing storage peripherals: the robots present, the different drives dedicated to Arcsys, etc.
- Managing media available for robots, with the media's id, barcode, capacity, remaining space, status, state, the drives that can be accessed, etc.
- Location of wrappers on the media, with the associated media and its placement (in offset or file mark form) for each wrapper archived.
- Description of all jobs and actions performed for each job.
- List of errors per hardware component, such as the number of read/write errors on a given tape and the movement of an arm, etc.

ACSLS / SCSI

SCSI, like ACSLS, control the robot to which the controller is connected.

For Oracle robots, use ACSLS and SCSI. For robots from other providers, only the SCSI protocol is available.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Asynchronous Tasks

Introduction

The main asynchronous tasks running within the system are as follows:

- Archiving processes
- Restore processes
- System audit processes

Archiving Processes

Even though creating an archiving request is a synchronous task, the operation consisting of fulfilling this request is purely asynchronous.

Characteristics

Archiving tasks are performed in a time window. This window is managed externally, via a command telling Arcsys that it must start or stop all archiving Nonetheless, this window does not affect requests sent online nor the creation and acceptance of archiving.

Data is analyzed and reorganized to create archiving objects. Objects for archiving are sent to the media manager via an access interface. Archiving identifiers received from the media manager as well as the metadata are finally stored at the Arcsys relational database level.

Archiving request statuses are updated in the Arcsys relational database. Managing the statuses for each request creates synchronization points for the archiving. These statuses are used for Recovery/Restart procedures. In case of a programmed or unexpected system shutdown, the processes resume at the levels indicated by their status.

Data Organization

An archiving lot corresponds to each archiving request. This functional concept of a lot represents a set of objects that will be archived in the same archiving process and with the same indexing criteria.

A file called a "manifest" contains all indexing criteria allocated to the lot, as well as the description of all objects present in the lot, will be generated and delivered by the

 © Infotel 2014 - Copy forbidden without explicit authorization.

transfer server. This information contained in this (XML) file comprises the metadata associated with the lot.

On the storage media, the "Manifest" always precedes the TAR file containing the objects.

Figure 14. Organization of the Manifest File

Metadata

The metadata associated with the lot can be of two types:

- Structural metadata:

  Object name

  Type: The type of the archived object indicates the type of action that can be performed on the object (editor tool, structure analysis, converter, etc.) ;

  Size;

  Description (optional text describing the object)

  Associated structure: A file describing the structure of the object can be associated with it, if appropriate.

- Indexing metadata:

  This data is used to classify a lot (all its elements) with respect to a set of indexing criteria. The indexing criteria are thus comprised of a list of "Keyword/Value" pairs.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Request Statuses

There are nine different statuses that indicate the progress of the archive request. Apart from their informative value, they are necessary synchronization points in system recovery/restart.

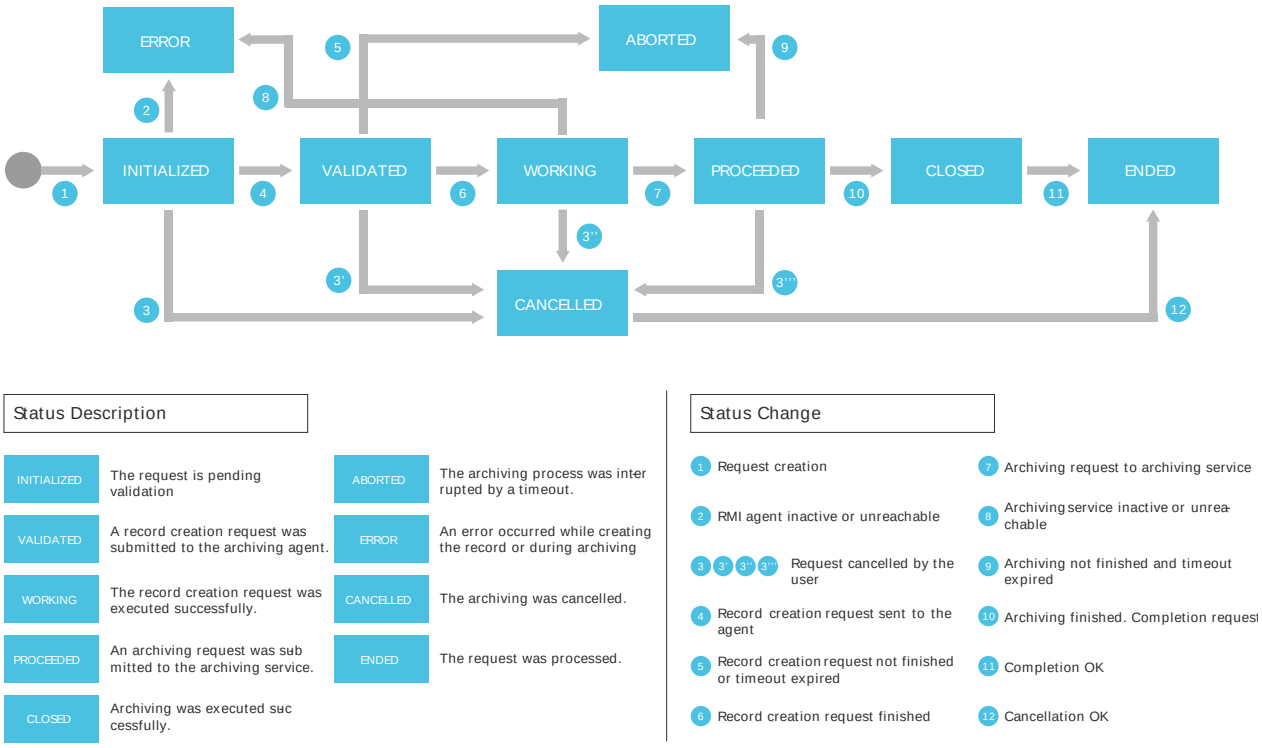

Figure 15. Process Flow of the Archiving Request Statuses

Restore Process

A restore process also exists.

Characteristics

It runs in an independent thread.

Restore tasks are performed in a time window. This window is managed externally via a command telling Arcsys that it must start or stop all archiving tasks.

The list of restore tasks is retrieved from the Arcsys relational database.

Request statuses are updated in the relational database.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Status management is used to create synchronization points used in the event of a Recovery/Restart.

Data Organization

All objects restored are transferred to the restore buffer sub-directory associated with the request. A file describing the restored objects will be generated by the engine and recorded in the same directory.

Request Statuses

There are nine different statuses that indicate the progress of the restore request. Apart from their informative value, they are necessary synchronization points in system recovery/restart.

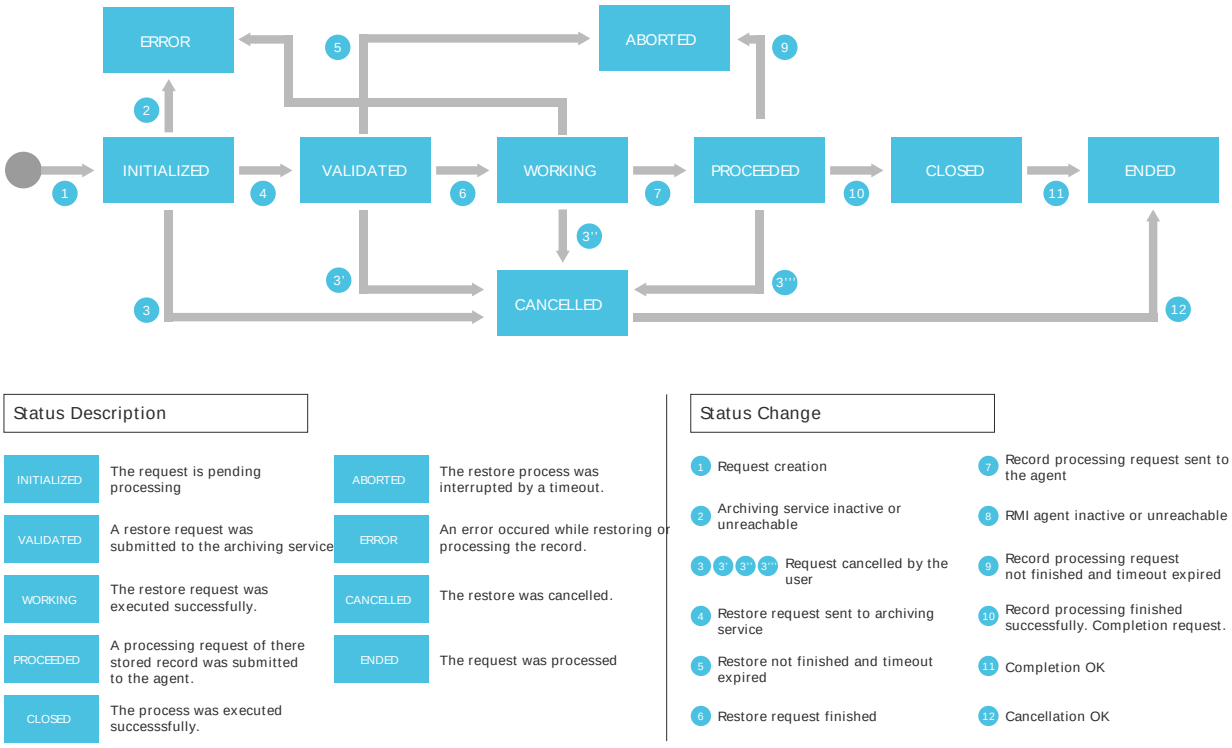

Figure 16. Process Flow of the Restore Request Statuses

Restore Modes

The restore process can be performed according to one of the following two modes:

- "Restore to client" mode

 © Infotel 2014 - Copy forbidden without explicit authorization.

In the "Restore to client" mode, the restore process will be executed until final availability of the data to restore on a specific remote space specified by the client. On starting the request, the client will have provided all the parameters required for connection and data transfer.

- "Restore to online" mode

In "Restore to online" mode, the restore process will run until the Arcsys engine is available on the online buffer of the Arcsys engine. The user must then download the data from the web interface.

System Check

The characteristics of the system check process are as follows:

- The system check is a permanent task that runs in an independent thread.
- It checks smooth operations of all components of the system and restarts faulty modules if necessary.
- It checks the use of work spaces.
- It closes sessions after detecting a certain period of inactivity.
- It purges "Close" or "Error" status requests according to retention periods defined for each of these cases.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Synchronous Tasks

Introduction

The main synchronous tasks running within the system are those generated by the engine or API accesses via their RMI interface. They can come in three types:

- Actions used for engine administration;
- Actions used to submit archive and restore requests;
- "Online" access to the record.

These synchronous tasks take place during user sessions. Hence, the execution of these tasks depends on the permissions associated with these sessions.

Engine Control

These are orders sent to the engine to stop, start, and query its current status.

The stop can be immediate or "graceful". A "graceful" stop awaits for the pending requests to be terminated before stopping the engine.

Request Management

Archive and restore requests take place via the Arcsys API. This management consists of a set of synchronous commands as seen below:

Creating a Request

This command initiates an archiving or restore process. Depending on the type of request created and the data transfer mode chosen, information concerning the location of the data (transfer protocol, service URL, access identifier etc.).

Cancelling a Request

All restore requests can be cancelled by the request initiator.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Status of a Request

All authorized users can query the status of a request (in archiving or restore).

"Online" Access to the Record

The advantage of this online record is immediate access to the elements it contains, which allows it to authorize synchronous access.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Advantages

 © Infotel 2014 - Copy forbidden without explicit authorization.

Introduction

Information is often managed in a company on different IT platforms, which is why Arcsys was built on a central architecture that can manage all the company's archiving, irrespective of environment type.

The Arcsys software is an archiving system that completely takes over issues of solution continuity and integrity:

- Arcsys captures information at source in real time, without penalizing IT production. In the same way, integrity checks are performed throughout the record retention phase, irrespective of the type of medium on which it is stored. Arcsys is probably the safest way to meet legal obligations for data retention.
- To guarantee system continuity, all records produced can be reread without the archiving software. In the same way, all internal structures are published so that system holders are completely independent and have free access to their records.

Above and beyond legal requirements, Arcsys is a software package used to archive all type of files and content for databases on the most common platforms, index record lots flexibly to enable searches based on company core business criteria.

The extreme scalability and independence from all types of hardware and software protects your investment and enables it to adapt to any increase in volume.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Advantages for Business Activity

Arcsys can meet legal obligations regarding data retention, but not just that. Integrity is a parameter set at the level of each object archived, meaning that Arcsys can be used both for legal obligations and for standard objects.

Depending on the company's business activity, data belonging to certain core businesses may be subject to archiving and require traceability.

Financial department managers may also be required to justify their business activity to fiscal authorities or even be required to comply with foreign regulations.

Secure storage and, in particular, the Arcsys classification methodology means you can find information quickly and the meet the claimant's demands.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Advantages for the User

Arcsys can be configured to completely automate periodic archiving operations. Nonetheless, users can still perform manual archiving operations via the GUI.

The archiving operation either makes a copy or moves objects. For the latter, they are deleted from their original position.

Metadata can be either acquired directly from the content of the object to archive (structure metadata) or defined via a definition structure (indexing metadata) to comply with regulations.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Advantages for the IT Department

Arcsys is a central archiving platform that hosts and consolidates all the company's archiving applications. Storage media can also be consolidated.

Arcsys manages the automatic transparent technological migration between storage media.

Arcsys rereads storage media periodically to prevent faults on the physical media.

The use of various storage media guarantees continuity of the archiving system.

The records generated are independent of all software components, so they can be viewed without having to stop the software that generated them.

With efficient administration tools, the archiving process is customized to meet all specific client needs.

The system is granular and scalable, which makes investments sustainable and reliable. This is particularly important to meet regulatory requirements and needs for long retention periods.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Compliance with Standards and Legislation

Arcsys takes into account standards issued from two different areas:

- Standards for secure preservation of digital data: NF Z42-013, ISO 14721 (OAIS)
- Standards for the management of electronic records: MoReq2, MoReq 2010

With respect to archiving, the NF Z 42-013 standard is the better known. This standard manages the integrity of a container: the storage medium.

French legislation (article 1316 of the code civil), however, requires the integrity of the content, from object creation to its presentation as evidence.

The difference between this standard and the law is therefore important, as before it is copied onto an electronic medium, in many cases the data has already existed on a simple magnetic medium. The search for integrity therefore covers the period preceding the copy (as close as possible to the original). In the same way, if the retention period exceeds the life expectancy of the medium, integrity must also cover the technological migration periods of the medium and cover the succession of storage media.

We can see that without questioning the spirit of the standard, strict integrity requires much more than simple compliance with the standard.

Arcsys offers a superior guarantee of integrity. The integrity certificate, calculated during the archiving procedure, will remain effective throughout the life cycle of the data, irrespective of the storage medium.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Appendix A. Appendices

 © Infotel 2014 - Copy forbidden without explicit authorization.

Glossary

Application Agent
There are two different types of agents at archiving level: application interface agents and user interface agents. An **application agent** can archive all the objects specific to an application (files, RDBMS table records, etc.), whereas a **web agent** performs both administration functions and manual archiving functions initiated by the user.

Aggregate (or Document Group)
An aggregate represents a logical group of objects to which it is possible to associate a number of properties (such as classification, retention period, disposal) inherited by all objects belonging to the aggregate. This is a concept used notably in MOREQ2010.

API *(Application Programming Interface)*
The APIs supplied by Arcsys are used by the product holder to completely customize a new Application or User interface to meet the precise ergonomic needs of a given application. A distinction must be made between:

- low-level APIs available in Arcsys Core, in RMI or SOAP protocol, providing access to most operations performed by Arcsys (administration, operation, archiving, searches, archive delivery);
- and high-level APIs, available in the ArcBWSoption. Web services in REST protocol provide access to a number of the most-frequently used methods.

Record
A record is an evidential document that is deemed sufficiently important by the creator to be managed by an ERM that will manage its life cycle (retention, disposal, etc.). A record represents an archived lot. A record is archived via a record request. Archiving a document creates a record.

Arcsys
ERM published by Infotel.Arcsys, misleadingly, refers to both the Arcsys Core product and all of its connectors and options.
See Also **ERMS**.

Authenticity
Through its authenticity, the document is proven to be what it purports to be, i.e. to have been created or sent by the person purported to have created or sent it, and created or sent at the time that it was purported (ISO 15489-1:2011 - 7.2.2).

Repository (previously base)
This is a logical entity in the Arcsys relational database. The company can define as many repositories as it wants, either to define a test set, to isolate an application, or for any other reason. These repositories are entirely independent of each other. They all have their own pattern and all have the same structure.

CLI *(Command Line Interface)*
A set of commands that deal mainly with archiving, delivery and searches. These commands can be implemented in scripts to call Arcsys services in application strings. These are also referred to as client batches.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Collection

Set of rules that a record must comply with. The collection is defined via the Web agent or Arcsys API, and comprises information contained in the database tables. A collection always refers to two rules: one concerning the '"**storage profile**" and that relating to the '"**indexing profile**". A collection is assigned to the record on the initial record request.
See Also **Storage profile (or storage policy)**, **Indexing mask or policy** *(or metadata template)*.

Archive Delivery

Archive delivery is an operation that makes a copy of a record available to a record requester. This term takes precedence over the term *restore*, which has another meaning at archiving level (restore in the sense of handing back the documents to the organization that created them or to its representatives, then destroying them). Archive delivery can be complete (misleadingly called a complete restore) or partial (*Partial Archive Delivery*, misleadingly called a partial restore).

Arcsys Connector

An Arcsys connector is an operational module generally requiring an additional license used to interface with an external software package (GED, ERP, Mail) for archiving and/or archive delivery to and from Arcsys. The main connectors are:

- ArcGED (with ArcGED for Nuxeo, ArcGED for Alfresco, ArcGED for Livelink, ArcGED for Documentum, ArcGED fro FileNet)
- ArcERP (with ArcERP for SAP)
- ArcMAIL (with ArcMAIL for Microsoft Exchange Server 2010, ArcMAIL for Lotus Notes)

Retention start date

Date from which a retention period must be taken into account. This is a concept used notably in MOREQ2010.

Disposal due date

This is the date at which the last retention date is reached. This is a concept used notably in MOREQ2010.

Destruction

This is an irreversible action that deletes the documents by applying disposal criteria. It can be associated with the retention of residual information in the relational database.

Proof folder

This is comprised of a record, a proof sheet, and additional items where required (functional traces, for example) that are used, by demonstrating the integrity and the authenticity of a document, for admission as proof.

Access rights

Access rights refer to the user profiles or groups authorized to access documents or sets of documents archived in the system. This concept is also referred to in Arcsys as **permissions**.

 © Infotel 2014 - Copy forbidden without explicit authorization.

Retention period

A duration expressed in days, months or years of object retention. This is a concept used notably in MOREQ2010.

Hash value

Also called an "integrity certificate" in cryptography, this is the digest of a message which guarantees a practically unique result that is impossible to reverse calculate. The most commonly used algorithms are MD5 (128 bits), SHA-1 (160 bits) and SHA256 (256 bits). The new versions of SHA2 also offer 384 and 512 bit hash values. Arcsys includes a module that is capable of dynamically calling several algorithms. The choice of an algorithm type is valid for all archived objects within the same Arcsys product version; compatibility with algorithms from the previous version is guaranteed. The associated term *hash function* is also used.

Wrapper or container (envelope)

Arcsys groups documents stored in the system in wrappers or envelopes, either created by Arcsys during the archiving process (in this case, files in TAR format), or created prior to Arcsys processing by the user or third-party processes (*native wrappers* in AFP or ZIP format, for example). The representation of a wrapper in the Arcsys relational database is called a *logical wrapper.* Its technical implementation is also called *MoverReference*. Last but not least, the representation of physical information stored in the wrapper in the optional ArcMover module is called *ArcMover wrapper information*. Its technical implementation is also called *MoverMedia*.

Export

An export procedure is used to retrieve the ERM object with a view to transferring to another system and if appropriate deleting it from the ERM. The export must include the object, its information stored in the relational database, its traces and ideally its management rules.

Disposal Hold

Arcsys can be used to "hold" one or more lots stored in the application. This puts a "hold" on the status of the lots by preventing any operation that could modify this status. Furthermore, the user is guaranteed that when the disposal hold is active, the lot will remain in the same state as it was at the moment it was put on hold. The action of removing the disposal hold is called a *release*.

Time stamping

Time stamping is a technique used to associate a document with a certain date in reference to a given and recognised time system. The date set in this way is an essential element for document authentication. Time stamping can be performed internally or by a third-party time stamp.

Integrity

The quality of a document that has not been subject to intentional or accidental destruction, alteration or modification.

Lot

Arcsys can group together several different objects that form a functional set in a client application in the same physical record. It is comprised of different types of objects: files, databases, or any other type of object managed by Arcsys. It is possible to deliver the entire lot or one of the objects contained in the lot. The MOREQ2010 record is translated in Arcsys implementation by a lot; the lot,

 © Infotel 2014 - Copy forbidden without explicit authorization.

as opposed to a MOREQ2010 record, can represent documents that are not yet archived.

Manifest

The manifest is an XML file that defines precisely the content of a record. The manifest contains: metadata associated with the record, structure metadata, a description of the physical files of records that follow, the object-by-object content of the record, object formats, object names, size, hash value, algorithm used to calculate the hash value, etc. This is a type of complete ID card for the record.

The manifest is always written on the storage media and precedes the record that it describes. This process is used to automatically describe the storage media (irrespective of the medium). With this system, users can understand media content and metadata without installing the software that generated the records.

MOREQ2010

MOREQ2010 (Modular Requirements for Records Systems) is a standard made available by the European Commission (DLM Forum Foundation) for electronic archiving. As Arcsys pre-dates MOREQ2010, it is taken into account in Arcsys Core via an add-on that is used to progressively include MOREQ2010 entities (aggregate, classification scheme, etc.), on top of the Arcsys entities (repository, collection, lot).

Arcsys Engine

The central archiving platform on which synchronous and asynchronous archiving, indexing and restore processes operate. The engine can spread threads over multiple processors. This guarantees dialogue and traceability between the agents that are associated to it.

Metadata element definition (or keyword)

Component of an indexing mask. We use the term "metadata element definition" rather than the term "keyword" as it is closer to MOREQ2010. The metadata element definition in particular defines the type of metadata (date, string, digital, controlled) and its input mask, for example.
See Also **Indexing mask or policy** *(or metadata template)*.

Nearline

Storage level managed through a media manager (XBSA interface). This storage level can be used with or without online storage. It creates copies on the storage media supported by the chosen Media Manager (disks, tapes, optical disks, etc.) This storage level is optional. All types of online storage, nearline disks or nearline cartridges support periodical content verification functions, continuity of the formats used and the portability of integrity when technical or planned migrations are performed via the collections.

Arcsys Core

The Arcsys Core represents all the "essential" Arcsys modules, i.e.

- Arcsys relational database
- Arcsys engine
- Arcsys application agent
- low-level Arcsys APIs (RMI and TCP/IP)
- Arcsys transfer server

 © Infotel 2014 - Copy forbidden without explicit authorization.

- Arcsys transfer service
- Arcsys Web agent

See Also **Arcsys**.

Object
The object is a basic archived unit that can be retrieved via Arcsys. Lots contain one or more objects. An object can be: a file, a directory, a table, a relational table, etc. The MOREQ2010 component is implemented by this object concept; the object, as opposed to a MOREQ2010 component, can represent a document that has not yet been archived.

Online
Storage level, which must be disk type, that makes records permanently available within an extremely reduced time period. Arcsys also allows use of a Nearline storage level (see definition).

Arcsys Option
Arcsys options are added to the Arcsys Core for additional functionalities. They do not necessarily require an additional architectural module. They may be subject to a separate license. The main options are:

- ArcAFP (AFP format management)
- ArcBATCH (batch actions on the database)
- ArcBWS (REST web services)
- ArcBI (decision-making module)
- ArcFSC (file system / database comparator)
- ArcMOVER (media manager managing file systems and robots in SCSI or ACSLS)
- ArcIP (record injection)
- ArcEP (record extractor)
- ArcPAK (record compression on ArcMover and native injection of compressed files)
- ArcRFT (full text search)
- ArcSIGN (internal and external digital signature management)
- ArcCFN (digital vault)

Classification Scheme
A classification scheme represents a hierarchical classification of activities in an organization used to classify archived objects and link them in this way to their creation context. A classification scheme is made up of categories organized in a hierarchy over a number of levels.

Disaster Recovery or Business Continuity Plan
A Disaster Recovery or Business Continuity Plan takes into account all measures required to ensure the continuity of business activities following an incident. This includes the recovery of business activities: the Business Recovery Plan (BRP),

 © Infotel 2014 - Copy forbidden without explicit authorization.

which provides the procedures required to resume activities immediately after the accident, is an essential part of the BCP. To facilitate implementation of the BCP, Arcsys provides procedures for site management, redundancy via replication of different storage zones and high availability of components.

Indexing mask or policy *(or metadata template)*
As is the case with storage profile, an indexing mask is a rule that is referenced by a collection. An indexing mask can be referenced by several collections. An indexing mask refers to the use of a set of Keyword = Value pairs. The keyword component is set to make sense in a specific business application (e.g. Accounting Day, Department, Account No., Account Holder, etc.). The value component can be either unrestricted, or restricted to a set of acceptable values (e.g. A, B or C), or in date format, or restricted by an input mask. Some pairs are defined as mandatory whereas others may be optional.

An application which uses an indexing mask through a collection must supply all Keyword=Value pairs as they are defined using this mask. Any indexing-related anomalies lead to the record being rejected for conformity. This record is then added to the list of records with anomalies.

The indexing mask is defined by an administrator via the Arcsys interface or APIs. It is comprised of a set of metadata element definitions.

Storage profile (or storage policy)
A storage profile is a rule that is referenced by a collection. The profile dictates the storage media which are successively implemented to hold a record, as well as the retention period for each media. The storage profile is defined through the graphical interface. Applications or business users use it indirectly through the reference to a collection. A storage profile can be changed over time to reflect new retention periods or new storage media. The policy covers storage units by logical pool.

Storage pool
The storage policy assigns a "zone" to a "profile". This is a logical storage pool, characterized in particular by its time period (e.g. 10 years).

Relational database
Essential component of the system, it contains all the data (excluding archived data) used by Arcsys for its operation. It includes logical entities called "repositories" (see definition).

Restore
This term is used misleadingly in Arcsys to refer to the concept of archive delivery. It is not accepted in archiving terminology as to mean transfer and then destruction.
See Also **Archive Delivery**.

Retention and disposal schedule
This comprises all the rules defining the record retention period for a company or an organization, according to risks of unavailability and information system access requirements. It specifies the disposal after these time periods.

ERMS
Electronic records management system: all the hardware, software and procedures that organize and manage the capture, retention and destruction, with a view

 © Infotel 2014 - Copy forbidden without explicit authorization.

to managing downtime risks and document over-conservation, as well as the availability of records in a company or organization.Arcsys is an ERM. *MCRS* (*MOREQ2010 Compliant Records System*) is also used to refer to an ERM that is compliant with MOREQ 2010.

Security
An ERM requirement that involves including documents whose use (confidentiality, risk of exposure) and/or integrity (non modification of content, non-alteration of media) should be closely monitored.

Disposal
This is the outcome of archived documents when the retention period ends, i.e. generally, destruction or transfer.
See Also **Destruction**, **Transfer**.

Deletion
MOREQ2010 provides the following definition for this concept: the act of deleting referential data so that no trace remains. Generally speaking, an entity can only be deleted if is not used in stored record. Otherwise, it can only be destroyed and not deleted, thus leaving a residual entity.
See Also **Destruction**.

Tracking
This is the result of continuously creating, capturing and maintaining information about the movement and use of the system and objects (ISO 15489-1:2001, 3.19).

Transfer
In an archival sense, this operation sends an archived object to another IT system. Once the transfer is performed, the object can be removed from the ERM as needed. In OAIS terminology, a transfer represents more specifically the physical transmission of a record or a set of records by a service supplying an archive service. Not to be confused with the transfer of data in the purely technical sense, as with the Arcsys transfer server module.

Online delivery
This is an archive delivery that takes place in the form of a direct delivery of a document (for direct viewing or downloading) in a Web browser. This term takes priority over the term online viewing or, misleadingly, online restore.
See Also **Archive Delivery**.

Storage zone
The storage zone is a logical entity representing a physical storage space (set of file systems, robot, zoneNetBackup ,TSMorArcMOVERfor example).

Registered Trademarks

AIX and IBM are trademarks of IBM Corporation in the United States and/or other countries.

Firefox is a registered trademark of the Mozilla Foundation.

Linux is a registered trademark of Linus Torvalds.

 © Infotel 2014 - Copy forbidden without explicit authorization.

UNIX is a registered trademark of the Open Group.

Microsoft Windows, Windows NT, Windows Server, SQL Server, Internet Explorer are registered trademarks of Microsoft Corporation in the United States and/or other countries.

SAP is a registered trademark of SAP AG in Germany and other countries.

MySQL is a registered trademark of Oracle and/or its subsidiaries. MariaDB is a registered trademark of Monty Program Ab.

Java is a registered trademark of Oracle and/or its subsidiaries in the United States and other countries.

Infotel is a registered trademark of Infotel SA.

UNIX is a registered trademark of the Open Group.

All other trademarks are the property of their respective owners.

We want to know your opinion!

For all comments on this manual, please contact Arcsys technical support on https://techsupport.infotel.com/, by email (techsupport@infotel.com) or by telephone (+33 (0)1 48 97 59 61) specifying the reference of this document.